DOCUMENT RESUME

ED 473 291 JC 030 141

TITLE	Partnership for Excellence (PFE) Evaluation Project: Summary of Peer Review Teams for PFE Evaluation Project.
INSTITUTION	RP Group of California Community Colleges, Santa Ana.
PUB DATE	2001-02-00
NOTE	13p.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Accountability; Community Colleges; *Evaluation Criteria; *Management by Objectives; Outcome Based Education; Performance; *Performance Based Assessment; Performance Factors; *Research Methodology; Two Year Colleges
IDENTIFIERS	*California Community Colleges

ABSTRACT

The Research and Planning (RP) Group for the California Community Colleges (CCC) supports the Partnership for Excellence (PFE) effort to improve student success as measured by the goals, yet the RP Group sees the goals as limited in scope regarding the full work of the CCC. In addition, the RP Group is realistic about the political forces at work that do not necessarily find the system's efforts and its own appraisal of results credible. This document outlines what the RP Group perceives as risks of inappropriate assignment of community colleges to a position in a ranking system: (1) over-complexity and oversimplification; (2) premature measurement; (3) misidentification of performance outliers; (4) harm to intended beneficiaries: students and communities; (5) potential for a significant error in fact that is likely to be interpreted as performance failure or deficiency; (6) assuming the general applies to the specific; (7) lack of robustness in the model and instability of measurement over time; and (8) interpretability and meaningfulness. The RP Group does not recommend abandoning the task at hand, but rather suggests that a series of longer-term steps be used to address the issue. Some of these steps include using cautionary language at the outset, refraining from premature application of rewards and punitive measures, and improving data sources and measures. (Author/NB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

RP Group for California Community Colleges
Partnership for Excellence (PFE) Evaluation Project
Summary of Peer Review Teams for PFE Evaluation Project

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. Luan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Project Report

February 2001

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 473 291

OVERALL COMMENTS

Major Cautions and Limitations of the Proposed Analytical Strategy:

The RP Group supports the Partnership for Excellence effort to improve student success as measured by the goals even as we realize that the goals are limited in scope regarding the full work of the California community colleges. We are also realistic about the political forces at work that do not necessarily find the system's efforts and its own appraisals of results credible. But we are also highly skeptical of a process that necessarily assigns some colleges to a "List of Shame" or even "List of Shining Stars" that result from "zero sum game" rankings in the name of partnerships and efforts to improve. Even with the intent to "level the playing field," which we heartily endorse, we do not believe that a sufficiently robust methodology can currently be derived to avoid the serious risks of inappropriate assignment to the top or bottom quartile of ranks.

- *The risks of over-complexity and oversimplification.*
  While the goals of the system are distributed proportionately among the colleges based on size of enrollment, their accomplishment stems from myriad conditions. The complexity of these considerations do not yield to the simultaneous need for a system simple enough to be easily interpreted yet robust enough to stand up to scrutiny. Such elegance depends on a tested theoretical model but as yet, such a tested theoretical (vs. exploratory) model to guide this process does not exist. On the other hand, over-specification of data promotes the illusion that all major factors have been included and fosters incredulity if all relationships and factors cannot be easily explained.

- *The risk of premature measurement.*
  There are serious problems with expectation of results from such recent interventions—especially longer timeframe outcomes such as transfer/degrees and cohort-based groups such as basic skills. Expectation of linear rises in outcomes and comparability across entities in those slopes, even all things being equal with respect to non-controllable factors, is not warranted and any good evaluation practice recognizes this. Yet the testing of measures to identify performance differentials assumes such early results. Lack of measurable results early on does not necessarily indicate lack of progress which takes some threshold of interacting factors to register.

IC030141

BEST COPY AVAILABLE

- *The risk of misidentification of performance outliers.*
  While rankings are common in "report card" methods, there is not necessarily a statistical difference between those found in one quartile and in another.[1]

- *The risk of harm to our intended beneficiaries: students and communities.*
  The status of our current tools, the lack of data sources for some good variables and poor quality of other data, and the lack of a tested theoretical model at this point in time makes the rush to identification of low and high performers dangerous. Serious consequences of identification of low performers are likely (beyond even the dollars involved, loss of bond measures and grants—such as reputation, institutional discouragement, or outflow of students). If identification is unwarranted or mistaken, grave disservice is done to our students and communities.

- *Potential for a significant error factor that is likely to be interpreted as performance failure or deficiency.*
  A solid framework of what we know and don't know is necessary to reliably determine a delineation of high and low performing districts. But besides the identifiable factors that can describe incoming students and other external circumstances (the non-controllable factors or adjustment variables), besides the differences in what works (the controllable institutional effects such as interventions and good practices applied distinctively for various groups of students), there is a great amount of inherent error (random effects, measurement error, misspecification error, lack of knowledge about other factors and relationships). The problem is, the extent of such error is not known and cannot be known at this point. Further, without even attempting to include institutional effects in the model, there is no means for determining the relative contributions of the error and the institutional effects.[2]

---

[1] Rankings imply there are actual differences among all ranking levels in the performance of institutions. That is, it implies that the #1 institution was better than the #2 institution which was better than the #3 institution and so on. Since we are using modeling and estimation procedures to develop the rankings, this is not true. There could be very large errors in estimating adjusted rates--particularly with the number of variables and low quality of data in the model. It may be the case that there are no statistically significant differences among the adjusted transfer rates of the institutions. The preliminary data from MIS showed that course completion rates of various colleges spread in very small ranges. One college's data showed that the difference is within 1%-2%. Ranking within such a small range is usually statistically insignificant. Yet, the Chancellor's Office will be calling some colleges "high" and some colleges "low" performers based on the quartiles in which their ranks fall. The rankings imply that there is a greater degree of precision in the process than there actually is.

[2] The only Adjustment Variables (AV) being considered are uncontrollable factors. The purpose for this is to ensure that the adjustment is based only on the effect of the "uncontrollable" factors on the OV. Thus, the residuals would reflect for each college, the actual change in percent with an expected change in percent if these "uncontrollable" factors were held constant (leveling the playing field). The rationale for doing this makes sense, however by ignoring the "controllable" factors, we are assuming that the effect of the "uncontrollable" factors on the OV is INDEPENDENT of the "controllable" factors. In other words, the OV variance accounted for by the "uncontrollable" factors is assumed to be unique. However, because "controllable" factors are not identified, we don't know to what extent this assumption is being violated. The model might identify some "controllable" factors and then perform partial correlations where the correlation between the AV's ("uncontrollable" factors) and the OV is computed while controlling for the newly identified "controllable" factors. Then use the partial correlation matrix as input for the regression. As a result, the parameter estimates for the AV's will be based on the unique effects of the AV's on the OV because the effects of the "controllable" factors would have been controlled for. This is a longer-term issue: to identify these factors as well. It is not realistic to do so at this point in time.

BEST COPY AVAILABLE

- *Assuming the general applies to the specific.*
  Any systems framework cannot (and should not) include all the unique non-controllable factors that indeed affect an individual college. The error per individual college is greater than within the system as a whole.[3]

- *Lack of robustness in the model and instability of measurement over time.*
  The debatable validity of the untested constructs and their relationships, the uncertain reliability of some of the data, and the fluctuation of some of the outcome data from year to year, indicate unstable measurement in the short term and immature modeling as of yet. The data reflect too high a sensitivity to unidentified factors—not necessarily controllable ones. What results is an approach that takes a scalpel to cut down a tree or a shotgun to incise a wound: the tools are not yet appropriate to the significance of the task.[4]

  At the RP Conference 2000, Patrick Terenzini (co-author of How College Affects Students: Findings and Insights from Twenty Years of Research) admitted to the lack of a well-rounded body of research for community college success. It is with that in mind that a number of major efforts are underway across the country to assess what is known about student success and especially powerful interventions and good practices that work well with our particular constituencies. Such efforts include our own Center for Student Success, Washington State's concerted study of best practices in their community college system, the national studies of student learning outcomes like those sponsored by Pew Charitable Trust, AACC, NPEC, NCTLA-National Study of Student Learning, and others.

---

[3] The proposal assumes that we can develop one model that will work for all the California Community Colleges—or at least that the key exogenous variables can be held constant to the extent that we can infer that the main differences among colleges are a result of their positive actions (or lack of them). In reality, the model coefficients, relationships among variables, and perhaps even the form of the relationships, will likely vary substantially among different student populations and college types and mission emphases. At a minimum, the model should separate the data by student type. Having separate models for traditional and nontraditional students (further work would be needed on how to define these categories) would greatly simplify the modeling task and help develop more accurate models within student groupings. Also, this issue could be tested empirically to see if separate models are needed. A test for the equality of coefficients between the traditional and nontraditional student models would enable determination of whether or not separate models are needed. (See also remarks in the transfer section.)

[4] Ideally a system of indicators measures distinct components of the system of interest, and also provides information about how the individual components work together to produce the overall effect. Indicators may be useful standard bearers for an institution when they are composed of contextual and process variables that suggest the environment for positive change that mediates policy inputs, student characteristics, institutional climate and socio-historical characteristics in building composite indicators of educational vitality in the community colleges. By focusing on the inputs, context, processes, and environments of the various community colleges, legislators and educators can seek to improve educational practice by suggesting how outcomes are affected by different educational policies, practices, environments, and contextual variables. In this way, indicators can begin to inform policymaking about what combinations of inputs, contexts, environments, produce observed outcomes. Such a system can be used to track and collect data on how colleges can act to improve learning outcomes for students. Although PFE, as originally developed, sought to focus on systemwide indices of quality assurance and eschew inter-institutional comparison, such institutional comparisons are becoming the norm. The model as currently proposed appears punitive with respect to reporting outcomes. We recommend a re-emphasis of the proposed model from focusing on the deleterious effects of being counted as a "low-performing" college, to a model that provides contextual information for college improvement. Such a re-emphasis would go much further in promoting one of the ostensible goals of accountability which is to improve college and student performance as well as make the model itself more robust and interpretable.

BEST COPY AVAILABLE

- *Interpretability and meaningfulness.*
  While indicators, like the dials on an instrument panel, tell us something about the institution, they do so in a way that reveals little about the what the indicator reflects, the context, and the environment. Thus their meaning is often unclear. Differences among institutions and the community of learners they serve, mean that, although two colleges may achieve the very same results on a certain indicator, their actual *accomplishments* might vary considerably. That each college/district contributes its proportionate share of system outcomes does not reflect individual mission emphasis nor how outcomes are affected by different educational policies, practices, environments, and contextual variables. However, to various publics, there is only failure or success, not nuance. This issue is of great concern to the RP Group as well as to the college community at large.[5]

In brief, what it needed is more theory-based modeling, fewer variables, and better methods to separate effects of "controllable variables" from error in the model. Exploratory testing of variables is prone to great error. But as with any good scientific and applied medical research, we build, challenge, change as we go while minimizing harm. Evaluation in this light can be either part of the improvement process or punitive in nature. While external directives point to the latter, we believe these initial steps could be used as incentive for reflection and improvement if managed carefully.

Primary Recommendation

We do ***NOT*** recommend abandoning the task at hand, but rather, *explicitly* identifying and evolving a series of longer-term steps to better address the issue. These include:

- Use of strong cautionary language at the outset and refraining from premature application of rewards and punitive measures
- Improved data sources and measures, such as:
  - Use of better databases, ones which include each college's contribution to individual's transfer/transfer preparedness, and improved cohort accuracy
  - Use of comparable basic skills levels (CSS project)
  - Finding a good measure of student's first language (not currently in our databases)
- Identification and testing of variable relationships, as proposed but also using different methodologies
- Expanding the variable type to include known effective interventions (controllable variables)
- Examining the potential of other approaches such as cluster analysis, student type as a simpler basis of analysis, and others.

---

[5] The reporting of unfavorable data in the absence of vital contextual information may actually have negative effects on improving the college efforts. It may also lead to a contradictory public response to this info rmation in the system as a whole. A stated goal of the Accountability system is to obtain funding from the public and the legislature by demonstrating the effectiveness of the colleges. As suggested by Oakes (1986, p. 30), this strategy may backfire. The public may view positive reports of college effectiveness as an indication that all is well with the system, and not approve additional funding beyond current levels. That is, good news may breed complacency. Or, if reports are issued that are critical of the college's effectiveness, the effect may be to punish the unaccountable colleges for malfeasance and vote to cut or limit funding. Public support may actually decline. Without presenting indicator data in a context that yields greater meaning, higher education institutions run the risk that policymakers and the public will misinterpret the data. Often, poorly thought-out policy is the result.

BEST COPY AVAIL

- Allowing for unique college factors, qualitative factors and alternative analyses by districts in the first round to test alternative models and other subtleties eluding the initial crude analysis. These analyses should, however, stand up to rigorous critique.

Methodology:
Several issues of overall methodology and types of variables are addressed briefly before commenting on approaches to individual goal areas.

Indicators are derived from student level variables. Assuming that all variables are available for each student after listwise deletion, how does this affect the composition of the final sample? In addition, although the model makes strenuous efforts to ensure random selection of cases (students) for analysis, in the final iteration of the model, the focus of the analysis shifts from students to colleges and districts. Thus the effects of randomization appear to have been lost. The model as proposed appears to have a built in bias toward colleges with larger enrollments. In the final analysis, the model does not so much "level the playing field" as much as it seems to construct the playing field along the lines of a community college with a large enrollment.

Also, it is difficult to judge if the model accurately portrays the actual distribution of student characteristics. The distribution of the continuous variables were not available for review by the panel. It is also not clear how missing data will be handled by the model, whether this will be done by mean substitution, or derived from known data about the students already collected to impute the missing value.

Sampling
Although the method described for selecting the student sample based on the last digit or two of the Student ID's would likely result in a systematic random sampling, this may not be the case depending upon the methods colleges use for assigning Student IDs. It would seem more straightforward and less subject to the vagaries of the data, to select a random sample of the data using the routines in a statistics package such as SAS or SPSS.

Regression analysis
The variable selection and coding procedures described appear to be bivariate in form, looking at the relationship between the outcome variable and one adjustment variable at a time. The results from this procedure could be quite different from a multivariate approach where all relevant variables are considered simultaneously. It would be preferable to have a theory guiding this process so we could specify the model producing the outcome. The model could then be tested as a whole rather than in bivariate pieces. Or, if a theory-based model is not available, we could use some multivariate variable selection methods such as backward elimination or stepwise selection of variables for the regression model.

Types of Variables
It is not clear which variables are MIS-based (the use and quality of which we are generally familiar with) and which are from external databases. There should be concern about how reliable and applicable to given colleges these external databases are. Researchers are familiar with many pitfalls in the use of specific measures and could be of further help with these as analysis of the model proceeds.

Some confusion among members of the review panels:
Understanding of the process and the models was not consistent across work groups. Most comments relate primarily to the following interpretation:

The model's objective is the adjustment of outcome measurements to "level the playing field." It is not proposed to be used for making predictions or establishing causality. Stage One searches for a set of adjustment variables (AV) that has some relationship to the outcome variable (OV) but are uncontrollable by the colleges and uses student data of each college (1% systematic random sampling) to calculate the value of each AV for each college. The model uses a sample of student enrollments (along with a sub-sample for cross-validation) called the "student file". The purpose of this sample is to identify potential adjustment variables (AV). Once identified (after numerous statistical tests), these AV's are redefined at the college level and used in a new data set called the "college file", which has 107 cases (one for each college).

In Stage Two the major statistical exercise for the model is to compute a linear multiple regression model based on the AV's. The data are presented in the format of a single table, each college representing one case. Each case has many variables and one aggregated value is calculated for each variable. The result of the regression (including a check on the adjusted R-square) is that a certain number of AV's will be selected to establish the linear regression model. These are put in blocks (using hierarchical regression to achieve this purpose) and the regression is run. The outcome of the regression provides a linear model with many fewer variables, say 4 or 5, defined as having power for affecting the OV.

A predicted OV is created for each college. Next, the percent change between the actual OV and the actual OV is computed. This is the "unadjusted" change. The "adjusted" change is computed as the percent change between the actual OV and the predicted OV.

In Stage Three the AV's are used to make adjustments on the districts' original OV if the college's measure on those predictive AV's defined by the linear model are in a disadvantaged or advantaged position (top or bottom quartile). Two ranks will be used to define the top and bottom performers on both lists. Stage Three combines the college OV's into district OV's. Then a decision will be made on which districts needs to go onto the contingency base. Multiple measures—raw and adjusted OV—are used because the latter lacks the validity necessary at this time to stand on its own though it is the preferable measure theoretically.

Some groups found the approach creative, the process comprehensive and detailed, the materials overall to be clearly written and well laid-out, the procedures described in the document very clear, the objectives in each of the steps well defined and the statistical methods to calculate adjusted and unadjusted rates residuals to identify low and high performing colleges sufficiently powerful. Others were confused by the explanation of the two rates, the nature of the evaluation vs. the statistical process, and especially the proposed "model." Whether the model was/should be exploratory or predictive or causal and the appropriateness of a theoretical vs. "search" strategy were debated. While some of this discussion is professional debate and the final product will be polished, it is also a clue that lay readers are likely to wrestle with the intricacies and meaning of the model as a basis for the findings.

SPECIFIC COMMENTS FOR EACH PFE GOAL

Transfer

Methodology

Given the large sample size, the question of the defensibility of the inferences made by the model merit further examination. With such a large sample, the probability that statistical significance will be found, and the variable will be allowed to remain in the model as a predictor is high despite relatively low correlation coefficients. Thus there is a potential problem with statistical versus practical significance. This problem is also generally linked to the problem of constructing a model using a statistical package rather than formulating a model based on a more careful mining of the literature on transfer and testing it.

The statement of the problem could benefit from some discussion of how to ***connect*** the relations found between certain input and process indicators and desired outcomes as shown by indicators. That is, if a relation is found between student characteristics and transfer rate or readiness, what is to be inferred? How will this study could be used to identify input and process factors that lead to observed differences in institutional transfer rates and proportion of students who are transfer prepared?

Determining if the measures used are adequate to address the research questions posed is hampered by the lack of clearly stated hypotheses derived from theory. The proposed model develops a method to categorize adjustment and other variables using a scheme which groups certain data as input, adjustment, and outcome. The use of these terms in categorizing these indicators is somewhat confusing since in the analysis they are not blocked or grouped according to any clearly explained rationale. In addition, because these data are thought to be indicators of quality, they all appear to be just as easily categorized as input variables for determining institutional transfer rates in one regression equation, and later grouped according to input and adjustment variables for explaining variance or the connections between college transfer rates and readiness. In this way the proposed model might be strengthened in terms of connecting the input-process-outcome variables to better understand the relation among these three sets of indicators.

Consideration should be given to grouping institutions for comparison purposes. This is done in several state Accountability systems. For example, in California, elementary and secondary schools are grouped according to characteristics of the students attending. Grouping variables include parent education and occupation, percent of limited English proficient students, mobility rates (students entering and leaving a school during the school year), and the percent receiving Aid to Families with Dependent Children (AFDC) (Kaagan & Coley, 1988). Thus, in the K-12 system, California has attempted to group different schools according to certain socio-economic and educational characteristics. This is in itself a recognition of the importance of context variables.

A major problem of inter-college indicators is that they are generally designed to assess the environment of the total college rather than the environment actually encountered by individual students within the college. An important, but often overlooked distinction particularly when dealing with institutions of tremendous variation in enrollment and size, is the importance of the

"sub-environments" of the college. Thus environmental proxies such as those proposed by the PFE model that purport to capture and describe the total institutional environment may not be useful because they will confound these subenvironmental differences. Therefore the model may need to gather additional data as to the particular environment encountered by the student, whether this is a college major, participation in a remedial program, members of a peer group, or the quality of effort demonstrated by the student. This may be available to the researchers conducting the study, or questionnaire data can be used and integrated with other contextual data to identify environmental factors that have some practical effect on the transfer outcome. Now the model shows a disturbing paucity of student experiential data that may have a highly significant relation to persistence and transfer.

Specific Variables

The model appears to use variables that are available to the MIS system, and does not seek to include variables on student behaviors, activities, course taking patterns and other data that have been shown to have a link to student transfer. The use of student intent as a predictor for example, versus a confirmation of whether the student actually behaves in a transfer-directed way (e.g., passes freshman composition or transfer level mathematics) suggest that the model might be constructed on the basis of what is available rather than what has been shown to work. The lack of student behavioral information suggests that the model may function statistically, but is a mis-specified or under-specified model. This is particularly true if college leaders are to use the transfer data as feedback for improving transfer outcomes at their college.

As suggested earlier, the model appears to have been created with scant reference to prior research in the area of student transfer. Rather than relying on a single measure to describe the transfer effectiveness of an institution, perhaps alternative measures of the transfer rate using nationally accepted and reported models can also be used. Several come to mind such as the transfer rate determined by Adelman, Cohen, Grubb, McIntyre, and NCHEMS. This would also help to educate the non-research audience as to the variability of the transfer rate depending on who is included in the denominator.

There is also some concern over the adequacy of transfer data. Although the data from public, in-state universities appear generally reliable, it is less certain that data from the private and independent colleges and universities will be complete. Also it should be stressed that the model will also miss many students who transfer to out of state institutions. In addition, the model imposes an artificial time constraint on the transfer of students that does not reflect the actual course taking patterns and completion rates of students (Adelman, 1992). Thus the actual number of students transferring will almost always be understated. Although we recognize this as unavoidable, it must be stressed to the various audiences for these data that the problem of under-reporting exists with respect to actually counting the transfers.

Degrees/Certificates

Methodology

Why is percent change in degrees and certificates awarded being used as the OV? The PFE goal is actual number awarded. If percent change were meant to allow for comparison of large schools to small schools, it would be better to use adjusted raw data than percent change. Using

percent change discounts previous efforts of those districts already doing well on this measure, even when adjusted for college size.

The model suggests that the raw OV data be adjusted by multiplying by the reciprocal of the rank in number of credit students. Given the goal, perhaps adjusting by rank in terms of number of degree and certificate seeking students would be more appropriate. If the OV is (as proposed) expressed in terms of a percentage change, some adjustment for college size is probably in order, but using reciprocal of rank (however determined) obscures relative size and overcompensates for the identified problem. Without going into any more detailed analysis, it is clear, that for the 100th ranked college, even a 100% raw score would be adjusted to only 1%. Even at the top, the second-ranked college would get an adjusted score equal to only one-half of its raw score while the first-ranked college (which may be only slightly larger) would retain its full percentage after adjustment. Multiplying by the reciprocal of percent of average college size (in terms of credit students, for example) would give a fairer adjustment, though even this is flawed. It may be necessary to rethink the entire question of adjusting a percentage for size of institution, since calculation of a percentage is itself a kind of relative size adjustment.

Specific variables

How will changes in the reporting of certificates of less than 18 units be accommodated?

Course Completion

Methodology

What is the rationale for using the percentage change from the previous year for the unadjusted rate rather than the current year's percentage? If the attempt is to capture improvement in successful course completion percentage, it may be better to use a longer time period to capture improvement. Measuring improvement with the regression coefficient is a more appropriate than the measure suggested.

The R-square for regression models should also be examined. This is a very important statistic to review when doing regression analysis. If the result is a statistically significant regression model with an R-square of .25, is it a powerful model?

It is a good idea to analyze course completion rates separately for transferable courses, vocational courses, and basic skills courses. Yet, the proposal assumes that one model can be developed for each type of course that will work for all the California Community Colleges. This assumption is unrealistic. The model coefficients, relationships among variables, and perhaps even the form of the relationships, will likely vary substantially among different student populations and college types. At a minimum, the model should separate the data by student type. Having separate models for traditional and nontraditional students (further discussion could be had about how to define these categories) would greatly simplify the modeling task and help develop more accurate models within student groupings.

There are concerns about using the ranking measure of the unadjusted completion rate difference between 1999-2000 and the immediate previous one year 1998-99. The data shows fluctuation over years. Based on experience, it is not practical to expect a straight linear upward increase of

the course completion rate each year. It may be better to rank the linear regression coefficient, slope of the trend between 1995-96 and 1999-2000.

Specific variables

The data sources are not clear. How are each of the factors in the model be measured? Some of the factors are hard to measure, such as "family economic history", "family obligations", "job market", "business demand for training", "alternatives to C.C. education", "language skill development", "access to transportation", "availability of public transportation". To be more specific, what specific data elements will be used for the statistical exercise? Take the "Draft Model Diagram" for an example, the diagram only describes a general concept or records a thinking process. The data variables and the source of data have not been clearly defined to factor those in the model. What specific MIS or other data elements will be factored in to the statistical exercise? How does the model defines 'academic preparation"? Although paragraphs 6 and 7 of 'the Draft Overview Comments" provides some explanation, it's not clear what data variables will be used to factor in the 23 rectangles in the diagram. How will the model factor in those variables–at the individual course level, or institutional level? If variables are to be included in the model, how are they to be measured? If they can't be measured, these factors shouldn't be included.

Workforce Development

Premise: Under PFE Goal 4, Workforce Development, what is being considered is numbers of students who successfully pass different levels of (SAM A) apprenticeships, (SAM B) advanced level vocational courses, and (SAM C) introductory vocational courses—not unlike, in many ways, Goal 3: Successful Course Completions (transfer, basic skills, and vocational). The California community colleges are required to target goals for increases of successful student completion of these three subgoals, and colleges have all had a chance to adjust ~~our~~ campus numbers based on ~~a~~ linear "predicted" improvement in student course taking success.

Methodology

The outcomes are based on numbers, not percentages of students who successfully complete vocational courses. In general, students successfully pass courses between 60% and 70% of the time. The system can be manipulated if a college increases the number of vocational courses offered, which is an avenue more likely to be followed by large colleges whose infrastructures can withstand low enrollment classes or who have extra funds to focus on expanding vocational enrollments—ergo "successful course completions."

There is a concern about using student/course level models to find predictors, when it is much different at a college level. This is too mechanistic, and possibly irrelevant. The model should use only college-level variables (i.e. college or student level variables aggregated to the college level), because the student level of analysis may not apply at the college level. Doing a two-step analysis, as proposed, moves up to the college-level without controlling for ~~the~~ variance at the student level. This can only be done in an HLM model, which is not recommended for ranking situations such as the current model. (Saying that women are more likely to succeed in vocational education classes within all colleges may not necessarily lead to saying that colleges with higher numbers of women in vocational education classes have higher numbers of success, because the aggregation/variance is different at each college. The model may as well just test the

latter at the college level.) Using the college level (size, vocational education class enrollments, local employment info, some economic level variable—e.g., income of closest city, some percent of gender, age, and ethnicities) is also the cleanest and clearest way to control for different types of colleges and locations. Colleges need to understand why certain variables are being used, and anything more complex will be alienating.

Factor analysis may show ~~us~~ which variables group together, but it must be kept simple for public understanding. Modelers might look at the correlations and try some OLS regression models on the less correlated ones, using theory to choose between them.

Specific variables

The most important uncontrollable variables—job shift change, home responsibilities, transportation, financial status, etc.—simply don't have readily available sources of data and data would be extraordinarily difficult to get ~~at~~. It also seems that this is a measure in which many of the factors are controllable variables—quality of instruction, availability of support services, scheduling, etc.

Concerning the list of 19 variables:
#2 & ~~#4~~—many vocational students take only a few units and wouldn't qualify for financial aid
#8—basic skills: the real issue is probably underprepared students who skip basic skill classes

The model should control for size of the college, for enrollment in vocational education courses compared to overall enrollment, and for the number of vocational education classes offered compared to offerings of all other courses. This would operationalize variable #4.

The model needs to control for urban/rural location of each college (as a proxy for access to local business/industry and the college's need to support vocational education /workforce preparedness of the student population). The CCCCO already has a nice matrix of size of college and population density, the latter of which may be a good proxy for urban/rural, but this does not predict unemployment rate. Different urban and rural areas may vary in employment opportunities, such as Santa Barbara vs. LA or Santa Cruz ~~rural~~ vs. Shasta. Maybe county unemployment rate would be better, as it would show job availability in a given area and maybe address the next item too. Population density and unemployment: since the Chancellor's Office already has the former, and the later is available by county from the state. Also, it is possible that the lower the unemployment rate, the fewer vocational education enrollments colleges will have. It is though this is true for overall enrollment, but not sure for vocational education enrollment specifically. Local unemployment rates, and/or population density can serve as a proxy for ~~the~~ demand for workers and availability of jobs (but it's not clear which direction for these variables predicts more enrollments).

The number of businesses within the college service area may be a valuable variable, but this data may be available in some regions and not in others.

The number of students declaring intent to complete certifications identifies interest in vocational education courses. However, this is very questionable data, because it is frequently registered by

students only on their initial application prior to any counseling and/or coursework. It is likely to change and/or be poorly thought out by students.

With regard to a variable identifying the proportion of the college budget dedicated to vocational education-type goals, there may be concern about VATEA funding varying among colleges in a way separate from enrollments.

Basic Skills

Methodology

The variable selection and coding procedures described are bivariate in form, looking at the relationship between the outcome variable and one adjustment variable at a time. The results from this procedure could be quite different from a multivariate approach where all relevant variables are considered simultaneously. It would be preferable to have a theory guiding this process so we could specify the model producing the outcome. The model could then be tested as a whole rather than in bivariate pieces. Or, if a theory-based model is not available, researchers could use multivariate variable selection methods such as backward elimination or stepwise selection of variables for the regression model. In general, it would seem that a much simpler approach to this would be to calculate rates separately for a category or categories of students. This is the approach used by the federal government in the Student Right to Know research.

Specific variables

The model should include specific institutional effects in the model, investigate factors of success in gateway courses, and use categories of students (e.g., traditional/non-traditional).

Questions:

Are all students to be included in the model?

Where will the data come from? How is it measured? Are the variables attached to each student in the sample or are they environmentally induced variables on the basis of some other second-order analysis?

Are appropriate external variables included (or only ones from MIS)?

Are there studies that show the relationship between the AV's and the OV?

IN CONCLUSION

The RP Group's Review Panels applaud the attempt to "level the playing field" if individual Districts are to be considered for evaluation purposes. The groups, in general, feel that the proposed solution would be good if the necessary data and theoretical support were available. However, the panels feel that the needed data and theory for this approach are not now available and will not be available in the short term. Attempting to construct the model and rankings in the current environment, it is feared, will be an overwhelming task and the rankings derived from the process will not be valid due to the concerns outlined above. The RP Group strongly encourages the Chancellor's Office to consider other, simpler, methods to achieve "leveling the playing field," and further, to emphasize constructive use of the evaluation data for improvement as a primary aim of the process. The RP Group would be pleased to assist should the Chancellor's Office decide to develop an alternative approach.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources information Center (ERIC)

NOTICE

Reproduction Basis

This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

This document is Federally-funded, *or* carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").